

## DOCUMENT RESUME

ED 478 325

SE 068 156

AUTHOR Tatsuoaka, Kikumi K.; Boodoo, Gwyneth M.  
TITLE Subgroup Differences on the GRE Quantitative Test Based on the Underlying Cognitive Processes and Knowledge.  
PUB DATE 2000-00-00  
NOTE 38p.; In: Kelly, Anthony E. and Richard Lesh, Handbook of Research Design in Mathematics and Science Education, Mahwah, NJ, Lawrence Erlbaum, 1999. p821-57.  
AVAILABLE FROM Lawrence Erlbaum Associates, Inc., 10 Industrial Avenue, Mahwah, NJ 07430. Web site: <http://www.erlbaum.com>.  
PUB TYPE Books (010) -- Reports - Research (143)  
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS  
DESCRIPTORS Curriculum Development; \*Educational Research; Ethnomathematics; \*Evaluation; Higher Education; \*Mathematics Education; Numeracy; \*Racial Differences; Research Design; Research Methodology; \*Sex Differences

## ABSTRACT

This chapter describes a study adopting the findings from an investigation of the cognitive processing and knowledge requirements of individuals taking the Graduate Record Examination Quantitative (GRE-Q) Test. The study investigates subgroup differences such as major field of study, gender, and racial-ethnic backgrounds of students taking the test. (KHR)

## 29

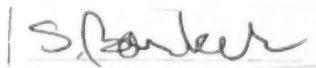
# Subgroup Differences on the GRE Quantitative Test Based on the Underlying Cognitive Processes and Knowledge

**Kikumi K. Tatsuoka**

**Gwyneth M. Boodoo**

*Educational Testing Service*

PERMISSION TO REPRODUCE AND  
DISSEMINATE THIS MATERIAL IN  
MICROFICHE ONLY HAS BEEN GRANTED BY



TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)

28

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

☐ This document has been reproduced as  
received from the person or organization  
originating it.

☐ Minor changes have been made to  
improve reproduction quality.

• Points of view or opinions stated in this  
document do not necessarily represent  
official OERI position or policy.

The cognitive processing and knowledge requirements of individuals taking the Graduate Record Examination Quantitative Test (GRE-Q) were investigated by K. K. Tatsuoka and Gallagher (in press). The study described in this chapter adopted their findings, especially the cognitive processing and knowledge requirements called "attributes," for investigating subgroup differences such as major field of study, gender, and racial-ethnic backgrounds of students taking the GRE-Q. The study also investigated the effects of content, contexts, and processing variables on students' performance on a GRE-Q. The taxonomy that divides the performance variables on mathematics tests into three categories—content, contexts, and processing variables—is based on Webb's (1984) suggestions. The content variables consist of the domains of arithmetic, elementary and intermediate algebra, and geometry, and the context variables are problem types, answer types, and the presentation of the tasks. The processing requirements are adapted from K. K. Tatsuoka and Gallagher in which several protocols were studied.

821

Tatsuoka's rule-space method (K. K. Tatsuoka 1983, 1985, 1995a) is employed for validating the cognitive processing model used in this study and for classifying the subjects on the GRE-Q into the students' latent states of knowledge. Several studies have demonstrated the usefulness of the rule-space model for such purposes (e.g., Birenbaum, Kelly, & K. K. Tatsuoka, 1993; K. K. Tatsuoka, 1990; K. K. Tatsuoka, Birenbaum, & Arnold, 1989; Tatsuoka, Linn, Tatsuoka, & Yamamoto, 1988; K. K. Tatsuoka, & M. M. Tatsuoka, 1992; M. M. Tatsuoka, & K. K. Tatsuoka, 1989).

This study provides information about the nature of the GRE-Q, namely, what the test measures, the underlying causes of group differences, and how test items can be improved to measure appropriate constructs for targeted subgroups. Several other researchers have investigated group differences too, but they used total scores and/or item scores (Bridgeman & Lewis, 1994; Wainer & Steinberg, 1992); no one has investigated the differences at the levels of the underlying cognitive processes and knowledge needed for a large-scale assessment. The results of the study provide information that will be helpful for developing new items appropriate for male or female GRE test-takers as well as Asian, Black, Hispanic, and White test takers, while measuring effectively the abilities required for graduate-level education in different disciplines. In addition, the results will enable test developers to design appropriate new items by helping them to decide which knowledge and cognitive skills ought to be included in new items or what types of items are better for measuring underlying cognitive skills.

Research on the association between scores on Scholastic Aptitude Test (SAT), the American College Test (ACT), or the GRE general tests and the quality of schooling or socioeconomic status of students indicate that curricular choices made during high school and undergraduate school correlate highly with the scores (Angoff & E. G. Johnson, 1988; Pennock-Roman, 1994). This study confirms these findings.

The chapter begins with a short account of the rule-space model used for the determination of the attribute probabilities for the examinees for Section I of the GRE-Q. After that the attributes identified for solving GRE-Q items are described, then the rule-space methodology is illustrated using the classification results of 10,000 examinees. The analyses of group differences by using the rule-space classification results are discussed next. These classification results consist of the weighted probabilities (the probabilities of using each attribute successfully) of the 14 attributes for an examinee who is classified into one of the

predetermined latent knowledge states. The chapter ends with the conclusions drawn from the study and a discussion of their implications.

### A BRIEF SUMMARY OF THE RULE-SPACE MODEL

The rule-space model is a probabilistic model for cognitive diagnosis that offers a general and cost-effective method. The mechanism of diagnosis used in the rule-space methodology is not domain-specific like other diagnostic systems (such as computer programs) using expert systems approaches. Thus, it is applicable to a wide range of content domains. Traditional paper-and-pencil tests as well as computerized tests can be analyzed by this approach at relatively low cost.

Diagnosing the cognitive errors of examinees taking the GRE-Q can be considered as a pattern classification problem that is designed to classify a sequential input of stimuli into one of several predetermined groups (Fukunaga, 1990; K. K. Tatsuoka, 1995a, 1997). For this study, the sequential inputs are the responses to the 30 items in Section I of the GRE-Q, and the predetermined groups are the various cognitive states of knowledge and capabilities resulting from the different degrees of knowledge or misconception in a quantitative domain. The rule-space model was developed to solve this classification problem. In the model, two main steps are essential for diagnosing cognitive errors: the formulation of a classification space and classification procedures, and the determination of classification groups. The classification groups are defined as students' knowledge states or cognitive states. Each knowledge state is expressed by an attribute pattern consisting of mastered and nonmastered attributes.

The formulation of a classification space is done by utilizing item-response theory (IRT) in order to preserve continuity with current psychometric theories (K. K. Tatsuoka, 1983, 1985, 1990; K. K. Tatsuoka & M. M. Tatsuoka, 1987). IRT probability functions are used as tools to formulate a Cartesian product space of IRT ability  $\theta$  and the variables,  $\zeta_s$ , which are sensitive to the unusualness of item-response patterns (Tatsuoka, 1984; K. K. Tatsuoka & Linn, 1983). Bayes' decision rules for minimum errors are used to classify a student into one of the predetermined knowledge states.

The classification groups (knowledge states) must be determined before the classification procedure is applied. The number of knowledge states can be several hundreds, or several

thousands (K. K. Tatsuoka, 1991). Several studies as, for example, an analysis of SAT mathematics (Buck, VanEssen, Tatsuoka, & Kostin, 1997; K. K. Tatsuoka & C. M. Tatsuoka, 1998a), have shown that even though 6,000 or 7,000 knowledge states exist, only 70 to 80 of them each contained at least 5% of the examinees, several hundred had only one or two students classified in them, and the remaining knowledge states were empty.

However, until we can design an experimental study in which unobservable cognitive tasks become measurable, it is impossible to observe knowledge and capability states directly. In practical testing situations, designing experimental studies for such data collection is not only unrealistic but also undoable. Because the rule-space model assumes that item scores are the only source of the scores obtainable, the model has to have the capability of making inferences about students' performances on unobservable cognitive tasks from observable item scores. The difficulty imposed on the rule-space model (namely, that only item scores are observable and that the underlying cognitive tasks are not observable) was solved by introducing a Boolean description function (BDF); (K. K. Tatsuoka, 1991; Varadi & Tatsuoka, 1989). The BDF makes the model applicable to many domains of interest. Unobservable cognitive states are determined systematically by applying various theories in Boolean algebra to item data. The BDF maps these cognitive states into observable item score patterns (called ideal item-response patterns).

The basic idea that initiated an application of Boolean algebra to the determination of knowledge states originated from a matrix representation of the relationships between attributes and items. The relationship between the cognitive tasks (attributes) identified and the items is expressed by an incidence matrix  $Q$ , whose order is  $K \times n$ , where  $K$  is the number of cognitive tasks and  $n$  is the number of items. If item  $j$  involves cognitive task  $k$ , then  $q_{kj} = 1$ , otherwise  $q_{kj} = 0$ . Thus, each item can be characterized by a set of cognitive tasks required for solving it. The set of subsets of row vectors (corresponding to attributes) forms a Boolean algebra. Therefore, all possible item-response patterns that are associated with attribute-mastery patterns can be derived from this incidence matrix  $Q$ . An example of an incidence matrix, which is described later, is given in Table 29.1.

## OBTAINING ATTRIBUTE PROBABILITIES USING THE RULE-SPACE MODEL

### The Data Set

The data used in this study are the item responses of 10,000 examinees who took the GRE-Q in 1991. The first 30 items in the Quantitative Test (Section I) were used for our analysis. Examinees used in the analyses were majoring in 33 areas of study from six categories: humanities and art, natural sciences, physical sciences, engineering and nonengineering, social sciences, business, and social work and education (see Appendix A for the fields studied). Differences in the use of the attributes demonstrated by students in the 33 major fields of studies, as well as differences in their use by gender and race-ethnicity are investigated in terms of the 14 attributes described in the article by K. K. Tatsuoka and Gallagher (in press) and given in Appendix B.

### Phase I: Obtaining an Incidence Matrix for Use in Rule-Space Analysis

An attribute for a task on the GRE Quantitative Test is a description of the content, processes, and contexts that an examinee must know in order to complete the task successfully.

#### *Identifying the Attributes*

As noted earlier, the attributes (the cognitive processing and the knowledge) required to complete the tasks specified in Section I of the GRE-Q were investigated by K. K. Tatsuoka and Gallagher (in press) and are listed in Appendix B. These researchers identified the four content, two context, and eight processing variables (attributes) needed to complete the tasks measured by the GRE-Q successfully. Readers are referred to the K. K. Tatsuoka and Gallagher article for a full description of the procedure followed in identifying the 14 attributes.

#### *Developing the Incidence Matrix*

The 30 items from Section I of the GRE-Q were coded by the 14 attributes in a Q matrix, summarized in Table 29.1, in which  $q_{kj} = 1$  if attribute  $k$  is involved in item  $j$ ,  $q_{kj} = 0$  if not (K. K. Tatsuoka, 1990).

*Validating the Incidence Matrix*

Regression analysis was used to regress the item difficulties on the attribute vectors. As mentioned in K. K. Tatsuoka and Gallagher (in press), the 14 attributes explained 78% of the variance in the 30 item difficulties. The process attributes, especially  $A_{10}$ ,  $A_{11}$ ,  $A_{13}$ , and  $A_{14}$  are strong predictors of item difficulty. Difficult items usually involve quite a few such process variables, and easy items do not. The same result was obtained for SAT mathematics (K. K. Tatsuoka, Birenbaum, Lewis, & Sheehan, 1993).

TABLE 29.1  
Incidence Matrix for the 14 Attributes Required to Answer Correctly 30 Items on the  
GRE Quantitative Test

| <i>Items</i> | <i>Difficulty of Item</i> | <i>Attributes Involved in Items</i> |
|--------------|---------------------------|-------------------------------------|
| 1            | -.995                     | 1, 6, 7                             |
| 2            | -1.362                    | 4, 6, 9, 12                         |
| 3            | -.808                     | 2, 6, 8                             |
| 4            | -1.023                    | 2, 5, 6, 7                          |
| 5            | .077                      | 4, 6, 9, 11, 14                     |
| 6            | -.634                     | 1, 6, 7, 9                          |
| 7            | .097                      | 4, 6, 7, 8, 10                      |
| 8            | .584                      | 2, 3, 6, 7, 8, 10, 13               |
| 9            | .398                      | 4, 6, 10, 11, 13                    |
| 10           | 1.207                     | 3, 6, 7, 8, 11, 12, 14              |
| 11           | 1.362                     | 1, 2, 6, 7, 10, 11, 12, 14          |
| 12           | -.086                     | 5, 6, 7                             |
| 13           | .069                      | 3, 6, 7, 8                          |
| 14           | .785                      | 1, 6, 7, 10, 11                     |
| 15           | 1.351                     | 1, 4, 6, 9, 10, 12, 13, 14          |
| 16           | -1.389                    | 1, 5, 6                             |
| 17           | -.609                     | 3, 8, 9                             |
| 18           | -.949                     | 2, 8                                |

(continued)

## 29. SUBGROUP DIFFERENCES ON THE GRE 827

|    |        |                       |
|----|--------|-----------------------|
| 19 | .039   | 1, 7, 9, 11, 12       |
| 20 | .354   | 4, 8, 9, 10, 11, 13   |
| 21 | -2.475 | 1, 12                 |
| 22 | -.719  | 1, 12, 14             |
| 23 | -.897  | 1, 7, 12              |
| 24 | .401   | 1, 5, 7, 12           |
| 25 | .819   | 1, 5, 7, 10, 12       |
| 26 | .312   | 4, 7, 8, 12, 14       |
| 27 | 1.247  | 1, 2, 5, 7, 8, 12, 14 |
| 28 | 1.168  | 3, 8, 9, 13           |
| 29 | 1.294  | 4, 5, 7, 8, 10, 12    |
| 30 | 2.580  | 1, 3, 7, 10, 11, 14   |

TABLE 29.2

Summary of the Weights Associated With the 14 Attributes in the Regression Analysis and the Attributes Required to Answer Items Correctly

| Attribute | Coefficient in Regression Probability <sup>1</sup> | Items Involving Attribute               |
|-----------|----------------------------------------------------|-----------------------------------------|
| 1         | 0.12 0.78                                          | 1, 6, 11, 14, 15, 16, 19, 21-25, 27, 30 |
| 2         | -0.27 0.49                                         | 3, 4, 8, 11, 18, 27                     |
| 3         | 0.19 0.71                                          | 8, 10, 13, 17, 28, 30                   |
| 4         | -0.33 0.61                                         | 2, 5, 7, 9, 15, 20, 26, 29              |
| 5         | 0.67 0.29                                          | 4, 12, 16, 24, 25, 27, 29               |
| 6         | -0.23 0.25                                         | 1-16                                    |
| 7         | 0.71 0.01                                          | 1, 4, 6-8, 10-14, 19, 23-27, 29, 30     |
| 8         | 0.59 0.13                                          | 3, 7, 8, 10, 13, 17, 18, 20, 26-29      |
| 9         | 0.49 0.09                                          | 2, 5, 6, 15, 17, 19, 20, 28             |
| 10        | 0.81 0.03                                          | 7-9, 11, 14, 15, 20, 25, 29, 30         |
| 11        | 0.78 0.02                                          | 5, 9-11, 14, 19, 20, 30                 |
| 12        | -0.28 0.30                                         | 2, 10, 11, 15, 19, 21-27, 29            |
| 13        | 0.93 0.03                                          | 8, 9, 15, 20, 28                        |
| 14        | 1.06 0.00                                          | 5, 10, 11, 15, 22, 26, 27, 30           |

<sup>1</sup>R<sup>2</sup> = .89, adjusted R<sup>2</sup> = .78

Table 29.2 gives the  $\beta$ -weights associated with each of the 14 attributes in the regression analysis and the items that require the successful use of the attribute for completion. Difficult items, such as Items 30 and 11, require the use of Process Attributes 10, 11, and 14 to obtain the correct answer, whereas easy items, such as Item 21, do not require the use of any of the process variables.

## Phase 2: Rule-Space Analysis

### *Determining Ideal Item-Response Patterns, Latent Knowledge States, Latent Classes*

When we have obtained an acceptable incidence matrix, we are ready for Phase 2. Before the classification procedure can start, we have to determine and prepare the classification groups (the latent knowledge states or latent classes).

For a set of items, the ideal item-response patterns correspond to examinees' possible knowledge states identified by the attributes required to complete the items successfully. For example, for the 30 GRE items, there are  $2^{30}$  possible item-response patterns ranging from incorrect responses to all of the items, resulting in the following pattern (0 = incorrect response, 1 = correct response):

|      |    |    |    |    |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|------|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Item | 12 | 34 | 56 | 78 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|      | 00 | 00 | 00 | 00 | 0 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |

to correct responses on some of the items as, for example:

|  |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|--|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | 11 | 11 | 01 | 00 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
|--|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

to a pattern of all correct responses:

|  |    |    |    |    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|--|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | 11 | 11 | 11 | 11 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|--|----|----|----|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

Not all of these  $2^{30}$  response patterns are ideal item-response patterns because they do not all correspond exactly to possible combinations of the 14 attributes representing examinees' knowledge states.

*An Illustration.* Let us suppose that each item measures a task that requires only one attribute for its correct solution, as follows:

|      |    |    |    |    |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|------|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Item | 12 | 34 | 56 | 78 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| Att. | 12 | 34 | 56 | 78 | 9 | 10 | 11 | 12 | 13 | 14 | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 1  | 2  |

There are many different knowledge states based on these 14 attributes. In fact, there could be up to  $2^{14}$  different knowledge states corresponding to all possible combinations of the attributes. For example, for examinees who can do  $A_1$  but not  $A_2$  through  $A_{14}$ , the ideal item-response pattern corresponding to this knowledge state is

1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0

In other words, these examinees should answer items 1, 15, and 29 correctly. For those who can do  $A_1$  and  $A_2$  but not  $A_3$  through  $A_{14}$ , the ideal item-response pattern is:

1 1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1

Similarly, we could derive the  $2^{14}$  ideal item-response patterns corresponding to each of the  $2^{14}$  attribute patterns.

*The GRE Case.* In fact, because each of the 30 GRE-Q items requires more than one attribute for a correct response (see Table 29.1), the combinatorial problem is more complicated than the one described previously, leading to a reduction in the number of knowledge states needed to explain test takers' performance on the GRE-Q. The BDF serves to reduce the number of knowledge states and hence the number of possible ideal item-response patterns. A computer program called BUGSHELL2, written by C. M. Tatsuoka, Varadi, and K. K. Tatsuoka (1992), derived 1,664 ideal response patterns based on the 14 attributes for classifying students across the 33 majors. The number of latent knowledge states in which at least one student was classified was 450. Sixty-six percent of the examinees were classified into one of the 1,664 predetermined latent knowledge states. The number of predetermined states, 1,664, was an enormously large number, and if there are too many classification groups in a space, the separation of such groups is very poor. Various techniques for reducing irrelevant states were used in BUGSHELL2, and hence the number of latent knowledge states was decreased to less than 1,000. Details of the classification procedures and technical issues were described in another report by K. K. Tatsuoka (in press), so readers who are interested in statistics should read that report. This chapter emphasizes the analyses of group differences.

*Classification*

Because all of the students do not respond to the 30 items in such a way that their responses are the same as the ideal item-response patterns corresponding to the predetermined latent knowledge states, a classification criterion is used for making a decision about how close each student's response pattern is to one of the centroids corresponding to the predetermined latent knowledge states in the rule-space methodology.

The classification criterion used was that the squared Mahalanobis distance (four dimensional rule-space was used for computing Mahalanobis distance) be less than 6, ( $D^2 < 6$ ); that is, if the squared Mahalanobis distance between student X and the centroid of a distribution representing a latent knowledge state is less than 6 ( $1 - p < .80, df = 4$ ), then X is judged to be classified. The probability of misclassification (error probability) and the posterior probability of X coming from the group in which his or her latent knowledge state is classified are computed. The classification rate is calculated by taking the ratio of the classified examinees to the total number of examinees in a sample group (; C. M. Tatsuoka & K. K. Tatsuoka, in press; K. K. Tatsuoka & M. M. Tatsuoka, 1987). Table 29.3 summarizes the number of students in each major, the classification rates, and the average probabilities of misclassification in each major.

TABLE 29.3  
Summary of the Number of Students in Each Major, Classification Rates, and the  
Average Probabilities of Misclassification in Each Major

| Major        | Number of Students <sup>2</sup> | Classification Rate | Average Probability of Misclassification |
|--------------|---------------------------------|---------------------|------------------------------------------|
| Agriculture  | 172                             | .62                 | .26                                      |
| Anthropology | 109                             | .59                 | .25                                      |
| Architecture | 102                             | .73                 | .25                                      |
| Art History  | 99                              | .67                 | .28                                      |
| Biology      | 796                             | .62                 | .22                                      |

(continued)

<sup>2</sup>Total  $N < 10,000$  because some major groups were small and were omitted from the analysis.

## 29. SUBGROUP DIFFERENCES ON THE GRE 831

|                                           |       |     |     |
|-------------------------------------------|-------|-----|-----|
| Business                                  | 168   | .68 | .30 |
| Administration                            |       |     |     |
| Chemical Engineering                      | 52    | .71 | .11 |
| Chemistry                                 | 177   | .65 | .17 |
| Child Education                           | 64    | .78 | .53 |
| Civil Engineering                         | 102   | .75 | .15 |
| Communication                             | 406   | .67 | .34 |
| Computer Science                          | 182   | .62 | .14 |
| Earth Atmospheric,<br>and Marine Sciences | 94    | .68 | .22 |
| Economics                                 | 164   | .68 | .15 |
| Electrical Engineering                    | 277   | .75 | .16 |
| Elementary Education                      | 347   | .67 | .40 |
| English                                   | 738   | .64 | .30 |
| Foreign Languages                         | 188   | .62 | .23 |
| History                                   | 433   | .63 | .27 |
| Home Economics                            | 63    | .78 | .48 |
| Industrial Engineering                    | 43    | .74 | .17 |
| Mathematics                               | 222   | .68 | .11 |
| Mechanical<br>Engineering                 | 185   | .69 | .12 |
| Medicine                                  | 708   | .65 | .34 |
| Performing Arts                           | 240   | .63 | .32 |
| Philosophy                                | 102   | .65 | .20 |
| Physics                                   | 138   | .74 | .12 |
| Political Science                         | 477   | .65 | .27 |
| Psychology                                | 1,381 | .64 | .25 |
| Secondary Education                       | 166   | .71 | .36 |
| Social Work                               | 76    | .71 | .49 |
| Sociology                                 | 225   | .64 | .50 |
| Special Education                         | 99    | .74 | .46 |
| Total                                     | 8,795 | .66 | .28 |

For agriculture majors, 107 students (see Table 29.3) were classified successfully, with a classification rate of 62% and an average error probability of .26. For electrical engineering majors, 208 students were classified, with a classification rate of 75% and an average misclassification probability of .16. The average misclassification probability for all students across the 33 majors is about .28.

If a less strict classification criterion had been used, say  $D^2 < 10$ , then more than 66% of the students would have been classified. However, because the purpose of this study was to investigate the differences between and among groups, it was important to preserve group characteristics by setting a strict classification criterion.

*Validity of the 14 Attributes: Variance of the IRT Ability Measure Accounted for by the 14 Attributes*

Three multiple regression analyses, regressing IRT  $\theta^3$  values onto the attribute-mastery patterns<sup>4</sup> (denoted by AMP), attribute-probability vectors<sup>5</sup> (APV), and item-response patterns (IRP) of all of the students classified were computed, and *R*-squares of .70 for AMP, .80 for APV, and .90 for IRP were obtained. Further, multiple *R*-squares and their adjusted *R*-squares for the 33 majors, female and male groups, Asian, Black, Hispanic, and White examinees were computed also; they are summarized in Table 29.4 and Table 29.5.

TABLE 29.4

The Use of Multiple *R*-Squares to Predict ability Levels By Major From Attribute-Mastery Patterns, Attribute-Probability Vectors, and Item-Response Patterns for the GRE-Q

|          | Humanities and Art    |             |          |                   |          |                 |            |
|----------|-----------------------|-------------|----------|-------------------|----------|-----------------|------------|
|          | Architecture          | Art History | English  | Foreign Languages | History  | Performing Arts | Philosophy |
| <i>N</i> | 74                    | 66          | 470      | 116               | 271      | 151             | 66         |
| AMP      | .79(.74) <sup>6</sup> | .85(.81)    | .81(.81) | .74(.70)          | .67(.66) | .76(.74)        | .77(.72)   |
| APV      | .82(.78)              | .89(.86)    | .86(.85) | .78(.75)          | .78(.76) | .81(.79)        | .80(.74)   |
| IRP      | .91(.84)              | .99(.98)    | .97(.97) | .93(.91)          | .91(.90) | .93(.92)        | .91(.83)   |

(continued)

<sup>3</sup> The values are estimated from a two-parameter logistic model using item-response patterns.

<sup>4</sup> An attribute-mastery pattern (AMP) is a pattern of mastered and nonmastered attributes defining a knowledge state. A knowledge state can be defined by an ideal item-response pattern or an ideal attribute-mastery pattern.

<sup>5</sup> An attribute probability vector (APV) is a vector of the estimated probabilities of the successful use of each attribute by an individual.

<sup>6</sup> Adjusted *R*-square.

## 29. SUBGROUP DIFFERENCES ON THE GRE 833

## Natural Sciences

|     | Agriculture | Biology  | Home Economics | Medicine |
|-----|-------------|----------|----------------|----------|
| N   | 106         | 490      | 49             | 459      |
| AMP | .79(.76)    | .71(.70) | .86(.80)       | .84(.83) |
| APV | .83(.80)    | .74(.73) | .93(.89)       | .89(.89) |
| IRP | .92(.88)    | .88(.88) | .99(.99)       | .99(.99) |

## Social Sciences and Business

|     | Anthropology | Business Admin. | Communication | Economics | Political Science | Psych.   | Sociology |
|-----|--------------|-----------------|---------------|-----------|-------------------|----------|-----------|
| N   | 64           | 114             | 271           | 112       | 309               | 883      | 144       |
| AMP | .88(.85)     | .83(.81)        | .84(.83)      | .78(.76)  | .66(.64)          | .72(.71) | .75(.73)  |
| APV | .91(.88)     | .88(.86)        | .90(.90)      | .79(.77)  | .76(.75)          | .80(.80) | .80(.78)  |
| IRP | .99(.98)     | .98(.97)        | .99(.99)      | .88(.83)  | .91(.90)          | .93(.92) | .92(.90)  |

## Social Work and Education

|     | Child Education | Elementary Education | Secondary Education | Special Education | Social Work |
|-----|-----------------|----------------------|---------------------|-------------------|-------------|
| N   | 50              | 232                  | 118                 | 73                | 54          |
| AMP | .84(.77)        | .82(.81)             | .89(.88)            | .83(.79)          | .90(.87)    |
| APV | .90(.86)        | .88(.88)             | .93(.92)            | .91(.89)          | .96(.94)    |
| IRP | .99(.99)        | .99(.99)             | .99(.99)            | .99(.99)          | .99(.99)    |

## Physical Sciences (Nonengineering)

|     | Chemistry | Computer Science | Earth, Atmospheric, Marine Sciences | Mathematics | Physics  |
|-----|-----------|------------------|-------------------------------------|-------------|----------|
| N   | 115       | 112              | 64                                  | 150         | 102      |
| AMP | .71(.67)  | .69(.65)         | .84(.81)                            | .67(.65)    | .68(.65) |
| APV | .77(.74)  | .72(.68)         | .89(.86)                            | .70(.68)    | .75(.72) |
| IRP | .85(.79)  | .83(.77)         | .98(.97)                            | .78(.74)    | .78(.69) |

## Physical Sciences (Engineering)

|     | Chemical Engineering | Civil Engineering | Electrical Engineering | Industrial Engineering | Mechanical Engineering |
|-----|----------------------|-------------------|------------------------|------------------------|------------------------|
| N   | 37                   | 76                | 209                    | 32                     | 127                    |
| AMP | .71(.67)             | .83(.81)          | .66(.64)               | .71(.63)               | .62(.59)               |
| APV | .74(.69)             | .86(.84)          | .71(.69)               | .77(.69)               | .74(.72)               |
| IRP | .87(.57)             | .94(.90)          | .81(.78)               | .96(.81)               | .81(.75)               |

Note: N = sample size; AMP = attribute-mastery patterns; APV = attribute-probability vectors; IRP = item-response pattern; GRE = Graduate Record Examination.

TABLE 29.5

The Use of Multiple *R*-Squares to Predict Ability Levels by Gender and Racial-Ethnic Group from Attribute-Mastery Patterns, Attribute-Probability Vectors, and Item-Response Patterns for the GRE-Q

|          |  | Gender                |          |
|----------|--|-----------------------|----------|
|          |  | Male                  | Female   |
| <i>N</i> |  | 2,647                 | 3,916    |
| AMP      |  | .68(.68) <sup>7</sup> | .76(.76) |
| APV      |  | .75(.75)              | .82(.82) |
| IRP      |  | .88(.87)              | .94(.94) |

|          |  | Racial-Ethnic Group |          |          |          |
|----------|--|---------------------|----------|----------|----------|
|          |  | Asian               | Black    | Hispanic | White    |
| <i>N</i> |  | 236                 | 369      | 219      | 2,819    |
| AMP      |  | .62(.60)            | .83(.82) | .79(.77) | .72(.72) |
| APV      |  | .74(.72)            | .90(.90) | .90(.90) | .78(.78) |
| IRP      |  | .84(.82)            | .99(.99) | .99(.99) | .91(.91) |

**Note:** Analyses were carried out to classify the 10,000 examinees separately by majors, and gender and race-ethnic groups. This accounts for the difference in the total numbers classified by gender and race in this table and by majors in Table 29.4. *N* = sample size; AMP = attribute-mastery patterns; APV = attribute-probability vectors; IRP = item-response pattern; GRE = Graduate Record Examination.

The item-response patterns predict IRT  $\theta$  very well for all of the groups because the  $\theta$ s are estimated from these item-response patterns. Across the 33 majors, the adjusted *R*-squared values vary from .99 for the majors in communication, education, home economics, and medicine (nurses are included) to .57 in chemical engineering. The average value for *R*-square

<sup>7</sup> Adjusted *R*-square.

for the 33 majors is .89. The attribute probability vectors, APV, account for about 80% percent of the variance of IRT  $\theta$ . This implies that the 14 attributes do almost as well as the 30 items in predicting IRT  $\theta$ . The APVs are better than the AMPs for predicting  $\theta$ . Similar trends were observed ( $IRP > APV > AMP$ ) for the two gender and the four racial-ethnic groups.

*Cross Validation.* The total sample of classified examinees was divided randomly into two subsamples, one of which was used as a training sample whereas the other functioned as a validation sample. A multiple regression equation using 14 APVs was estimated from the training sample, and the same regression equation was used to predict  $\theta$  values for the examinees in the validation sample. The correlation between the original and the predicted  $\theta$ s for the validation sample was .85.

## ANALYSIS OF THE RULE-SPACE CLASSIFICATION RESULTS

### Means and Standard Deviations of the 14 Attributes for the 33 Majors and 6 Gender and Racial-Ethnic Groups

*Differences Among Majors.* The means and standard deviations (SDs) of  $\theta$  and the 14 attributes across 33 majors, males, females, Asians, Blacks, Hispanics, and Whites are computed. The histograms of 14 attributes across various  $\theta$  levels for the 33 majors, genders, and racial ethnic groups are plotted in Figs. 29.1 through 29.5.

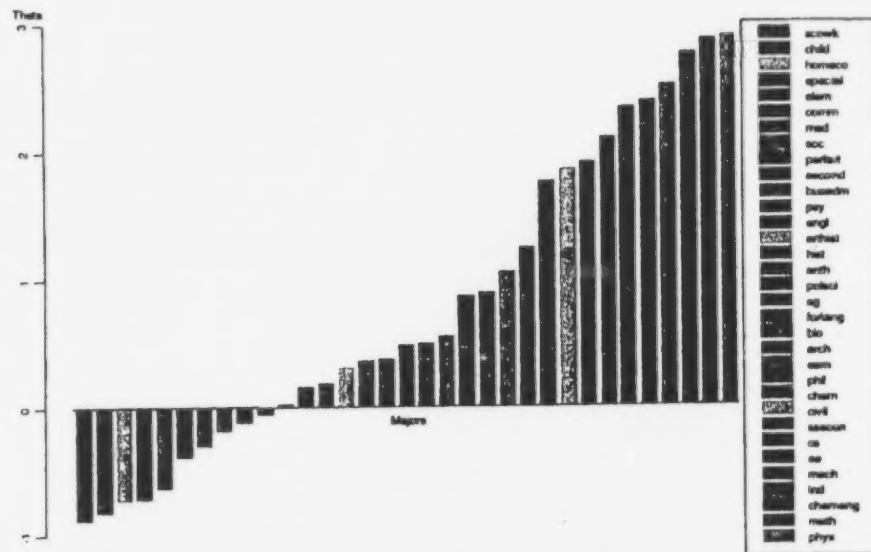


FIG. 29.1. Mean theta values for 33 majors.

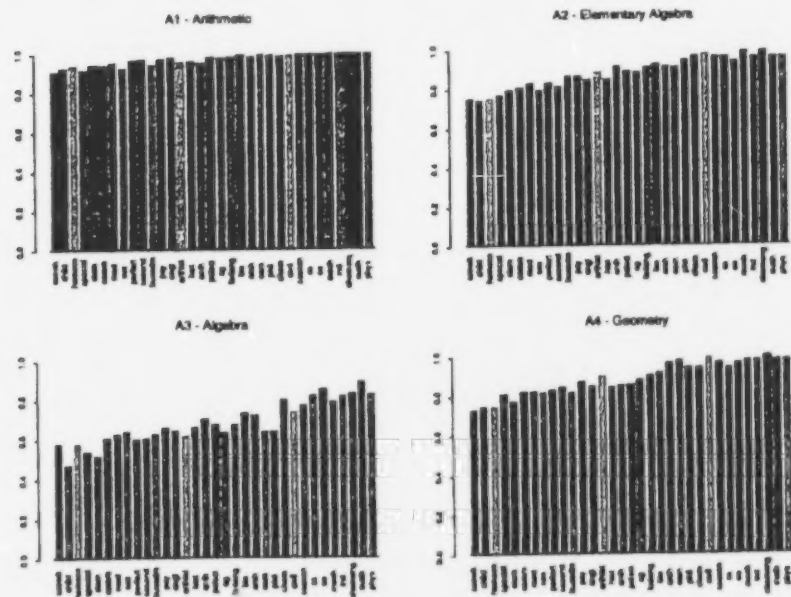


FIG. 29.2. Mean values of 33 majors for attributes 1 through 4.

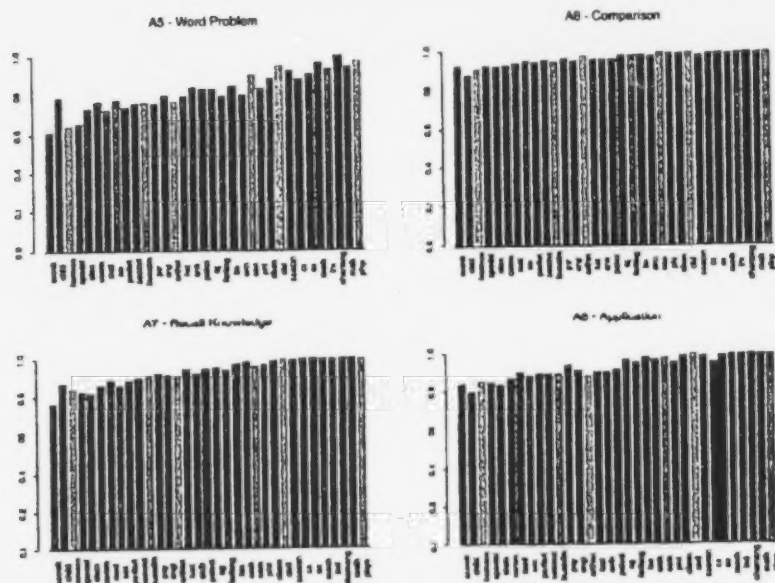


FIG. 29.3. Mean values of 33 majors for attributes 5 through 8.

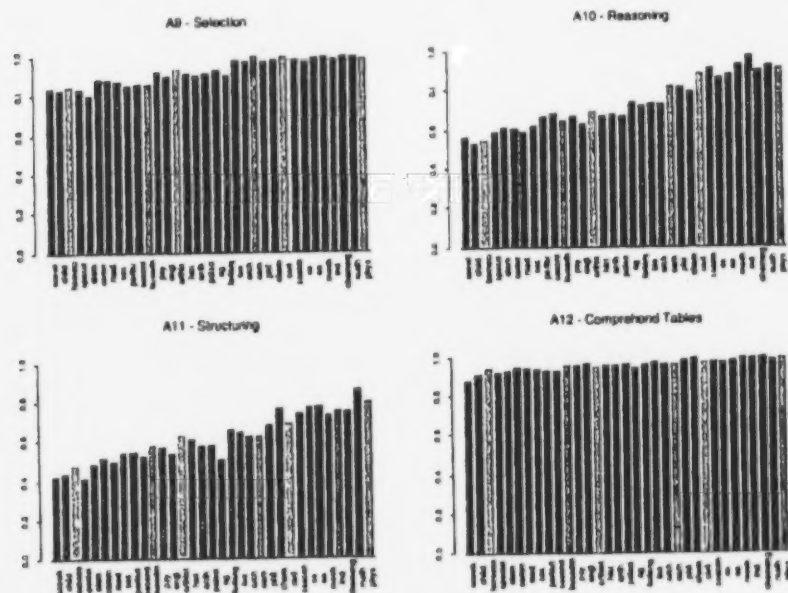


FIG. 29.4. Mean values of 33 majors for attributes 9 through 12.

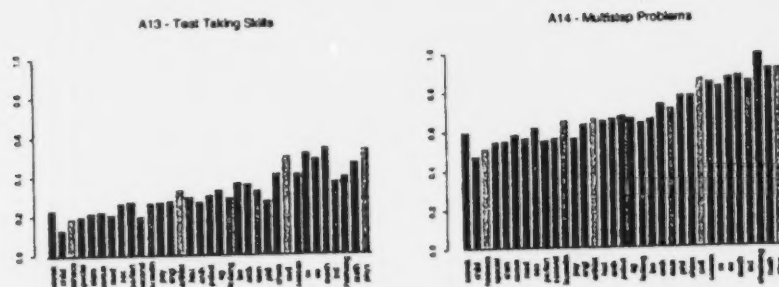


FIG. 29.5. Mean values of 33 majors for attributes 13 and 14.

FIG. 29.1 shows the bar graph of the mean IRT  $\theta$  values for the 33 majors. Social work majors have the lowest mean value for  $\theta$ , followed by child education and home economics majors. Physics and mathematics majors have the highest mean scores for  $\theta$ , followed by a group of engineering majors.

The average probability values and their standard deviations of all of the 33 majors for the 14 attributes are .97(.14) for Attribute 1 (arithmetic), .87(.26) for Attribute 2 (elementary algebra), .68(.37) for Attribute 3 (intermediate algebra), .87(.25) for Attribute 4 (geometry), .81(.35) for Attribute 5 (translation of word problems), .97(.14) for Attribute 6 (comparison item type), .93(.21) for Attribute 7 (application of knowledge), .93(.20) for Attribute 8 (application of rules and algorithms), .92(.20) for Attribute 9 (selection of rules and theorems), .71(.38) for Attribute 10 (reasoning and logical thinking), .59(.41) for Attribute 11 (analytical thinking and structuring the problem), .96(.15) for Attribute 12 (comprehend graphs, tables and figures), .31(.35) for Attribute 13 (test-taking skills), and .66(.40) for Attribute 14 (sorting multiple goals and solving them step by step).

The means of the 14 attribute-probability values of the examinees for the 33 majors are plotted in Figs. 29.2 through 29.5. In each figure, the x-axis represents the 33 majors ordered by their mean  $\theta$ -values. The common characteristic of the 14 plots is that the attribute probability values increase generally as the IRT  $\theta$  value increases although the slopes vary across the plots.

For the attributes that are the content variables, the slopes are gradual, not steep. The high means across all of the majors in Attribute 1 indicate that basic knowledge and skills in arithmetic are easy and it is mastered or nearly mastered by almost all of the groups. Attribute 3, basic knowledge and skills in intermediate algebra, is much harder than Attribute 2, which

is basic knowledge and skills in elementary algebra. Only physics, mathematics, chemistry, computer science, and the engineering (chemical engineering, industrial engineering, electrical engineering) majors achieved the probability of .8 in Attribute 3. Basic knowledge and skills in geometry (Attribute 4) are not difficult for most of the majors, except for those majoring in social work, child education, home economics, and elementary education.  $A_{12}$ , a context variable that includes following verbal instructions and comprehending terminologies, tables, and figures, seems easy for most majors.

As for the process variables, application attributes  $A_7$ ,  $A_8$ , and  $A_9$  have almost identical profiles across the majors. But the thinking-translation attributes are obviously more difficult, and their slopes are steeper than the application attributes. Attribute 13, test-taking skills, is the most difficult; indeed, it is more difficult than the reasoning skills,  $A_{10}$ , analytical thinking skills,  $A_{11}$ , dealing with multistep problems,  $A_{13}$  and translating verbal problems to algebraic, arithmetic expressions,  $A_5$ .

*Gender and Racial-Ethnic Differences.* With regards to gender differences, men achieve better than women in all of the attributes in the content and process areas and  $\theta$ . However, this result is expected because more men are majoring in science, engineering, and economics, whereas more women are studying education, home economics, and psychology. It confirms the known fact that the GRE-Q is much easier for mathematics, science, and engineering students. FIGS. 29.6 through 29.8 show the box plots of mean values for the gender groups.

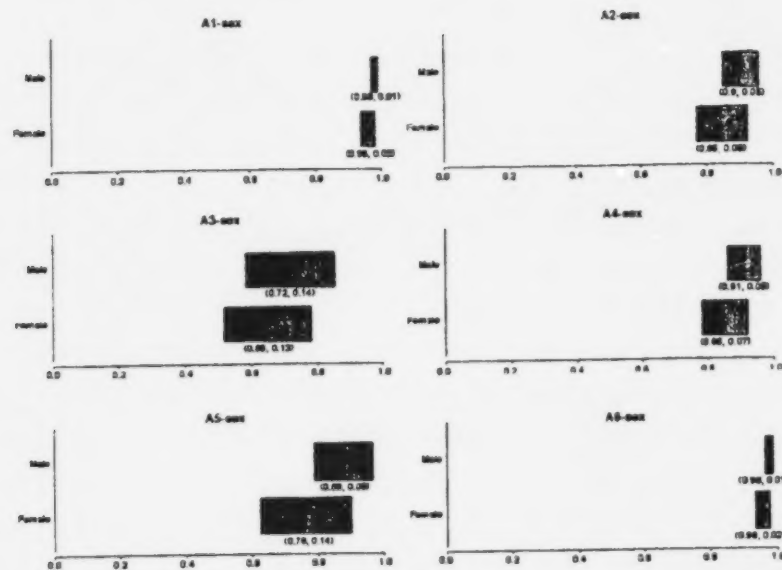


FIG. 29.6. Box plot of mean values for the two gender groups: Attributes 1 through 6.

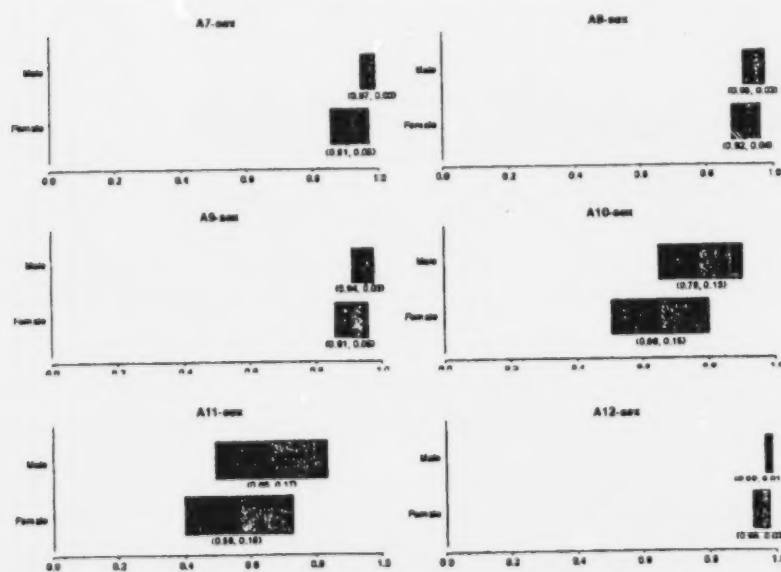


FIG. 29.7. Box plot of mean values for the two gender groups: Attributes 7 through 12.

BEST COPY AVAILABLE

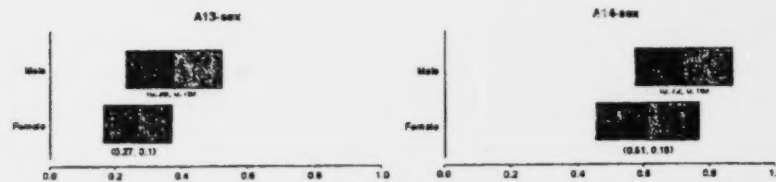


FIG. 29.8. Box plot of mean values for the two gender groups: Attributes 13 and 14.

The four racial-ethnic groups are also compared with respect to their attribute-probability scores (see Figs. 29.9, 29.10, and 29.11). Because the majority of Asians are majoring in mathematics, science, and engineering, this group outperformed the other three groups on the GRE-Q. Whites performed much better on the attributes in the content and context variables than Hispanics and Blacks, but the differences were much smaller in terms of their standard errors on the attributes in the process variables. The most difficult attribute,  $A_{13}$ , has the same mean value for Whites and Blacks; Hispanics' mean value is the smallest. However, Whites performed better than Hispanics, and Hispanics performed better than Blacks on the other attributes. In general, the differences in the attribute scores across the racial-ethnic groups are much smaller than the differences in  $\theta$  values.

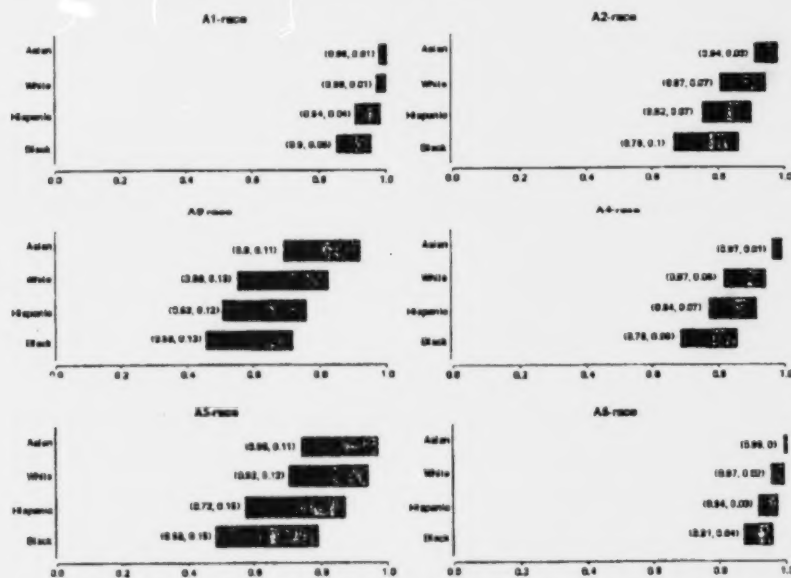


FIG. 29.9. Box plot of mean values for the four racial-ethnic groups: Attributes 1 through 6.

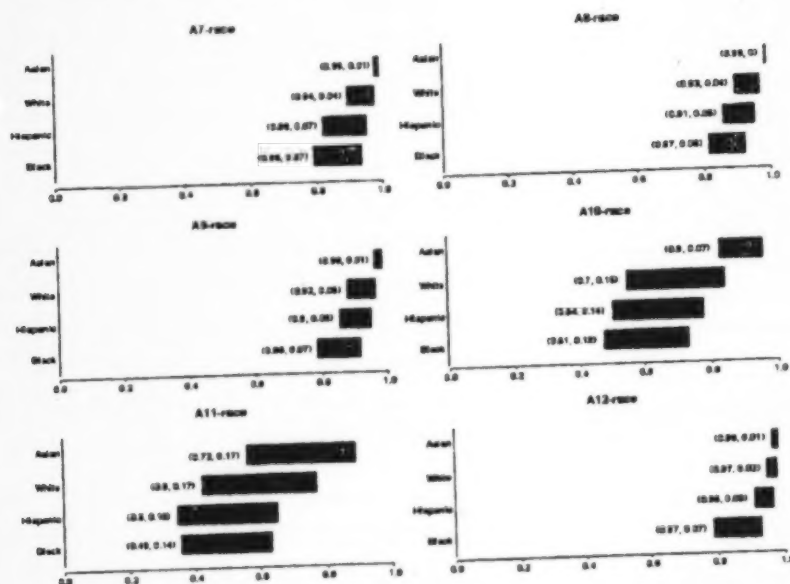


FIG. 29.10. Box plot of mean values for the four racial-ethnic groups: Attributes 7 through 12.

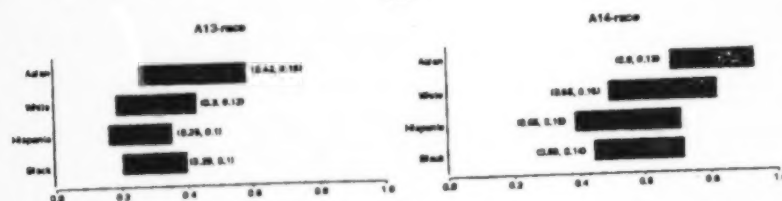


FIG. 29.11. Box plot of mean values for the four racial-ethnic groups: Attributes 13 and 14.

### Cluster Analysis

Cluster analysis is used for finding which majors, genders, and racial-ethnic groups formulate homogeneous categories with respect to the 14 attributes. The attribute scores are weighted probabilities whose values are between 0 and 1, so similarity measures are computed by Ward's (1963) minimum-variance method in which the distance is the analysis of variance sum of squares between two clusters added up for all 14 variables. The within-cluster sum of the squares is minimized over all of the partitions that can be obtained by merging two clusters at each step of cluster generation. The criterion for joining clusters at each level is to maximize

the likelihood by assuming a multivariate normal mixture, equal spherical covariance matrices, and equal sampling probabilities.

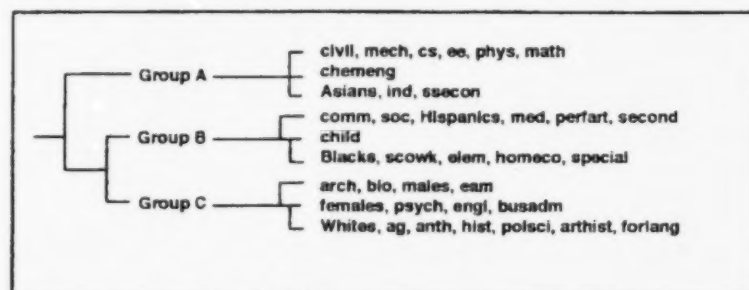


FIG. 29.12. Tree representation of clustered groups A, B, and C.

Using Ward's (1963) method, three major clusters were obtained and are depicted in FIG. 29.12. The groups included 31 out of 33 majors; chemistry and philosophy majors are omitted because they did not cluster with any of the three groups.

- Group A consists of industrial engineering, economics, chemical engineering, civil engineering, mechanical engineering, physics, computer science, electrical engineering, and mathematics majors.
- Group B is composed of social work, elementary education, home economics, special education, child education, communication, sociology, medicine, performing arts, and secondary education majors.
- Group C comprises agriculture, anthropology, history, political science, art history, foreign languages, business administration, English, psychology, architecture, biology, and earth atmospheric, marine majors.

The majors in science, mathematics, and engineering clustered in Group A. Asian students are in this group. The education majors clustered with social workers, medical students (mostly nurses), and sociologists in the second group. Hispanic and Black students are clustered in this group. The third group consists of most of the humanity majors, behavioral scientists, biology, and agriculture majors. White students are also in this group. Women and men also clustered in this group.

Each group has its unique characteristics. Group A consists of engineers and scientists, Group B is oriented toward services and education, and Group C consists of the humanities and

psychology. Asian students were classified in Group A, Black and Hispanic students in Group B, and White students in group C. In Group C, women cluster with psychology and English majors whereas men cluster with architecture and biology majors. Because the gender and racial-ethnic groups are not statistically independent of the majors, the discriminant analyses for the majors, gender groups, and racial-ethnic groups are carried out separately for the investigation of the subgroups' characteristics.

### Discriminant Analysis for the Majors

In order to enhance the characterization of the three Groups A, B, and C, a canonical discriminant analysis was performed with the 14 attribute scores as independent variables and the three groups as categorical dependent variables. The gender and racial-ethnic groups are analyzed independently. Group-level information, frequencies, and distances between the groups are listed in Table 29.6.

TABLE 29.6  
Squared Mahalanobis Distances Between Groups A, B, and C, and Their Frequencies

| <i>Groups</i> | <i>Frequency<sup>8</sup></i> | <i>Group A</i> | <i>Group B</i> |
|---------------|------------------------------|----------------|----------------|
| Group A       | 957                          |                |                |
| Group B       | 1601                         | 0.49           |                |
| Group C       | 3027                         | 1.49           | 3.36           |

Two canonical discriminant functions are obtained; the means of the three groups are given in Table 29.7 and plotted in FIG. 29.13.

---

<sup>8</sup>The three groups consist of 31 majors ( $N = 5,585$ ). They do not include the chemistry and philosophy majors ( $N=181$ ); these two majors did not cluster with any of the Groups A, B, and C.

TABLE 29.7.

Standardized Canonical Coefficients and Group Means for Groups A, B, and C

| Attributes                                           | Canonical Coefficient 1 | Canonical Coefficient 2 |
|------------------------------------------------------|-------------------------|-------------------------|
| A <sub>1</sub> (Arithmetic)                          | 0.077                   | 0.166                   |
| A <sub>2</sub> (Elementary algebra)                  | 0.088                   | 0.074                   |
| A <sub>3</sub> (Intermediate algebra)                | 0.243                   | -0.204                  |
| A <sub>4</sub> (Geometry)                            | 0.019                   | 0.062                   |
| A <sub>5</sub> (Word problem)                        | 0.277                   | -0.031                  |
| A <sub>6</sub> (Comparison format)                   | 0.107                   | 0.236                   |
| A <sub>7</sub> (Application of knowledge)            | 0.338                   | 0.446                   |
| A <sub>8</sub> (Application of rules and algorithms) | 0.100                   | 0.413                   |
| A <sub>9</sub> (Selection of rules and theorems)     | 0.126                   | 0.271                   |
| A <sub>10</sub> (Reasoning and logical thinking)     | 0.351                   | -0.477                  |
| A <sub>11</sub> (Analytical thinking)                | 0.272                   | -0.165                  |
| A <sub>12</sub> (Working with figures and tables)    | 0.067                   | 0.184                   |
| A <sub>13</sub> (Test-taking skills)                 | 0.267                   | -0.390                  |
| A <sub>14</sub> (Sorting subgoals)                   | 0.472                   | -0.397                  |
| Group A                                              | 1.170                   | -0.140                  |
| Group B                                              | -0.662                  | -0.155                  |
| Group C                                              | -0.020                  | 0.126                   |

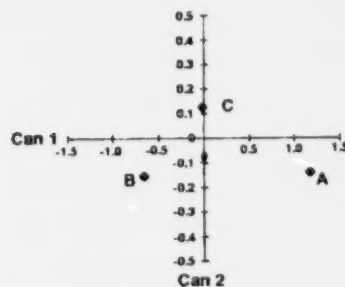


FIG. 29.13. Means plot of groups A, B, and C.

The larger the value of the coefficient, the more powerful the corresponding attribute is as a discriminator of the groups. The first canonical discriminant function indicates that the members of Group A can do well in "sorting and setting goals in multistep problems" (A<sub>14</sub>), are "knowledgeable in general mathematics at the high school levels" (A<sub>7</sub>), have strong "mathematical reasoning skills" (A<sub>10</sub>), can "translate word problems into algebraic or arithmetic relationships" (A<sub>5</sub>) well, can "separate a problem into components (implicit) and

restructure them into a solvable form" ( $A_{11}$ ), and can solve problems in unusual but clever ways ( $A_{13}$ ). As for the content areas, they know intermediate algebra ( $A_3$ ) well.

The second canonical discriminant function characterizes the students as being "knowledgeable in general mathematics at the high school level" ( $A_7$ ), able to "apply rules and algorithms for solving equations" ( $A_8$ ), and able to "select appropriate rules and theorems" ( $A_9$ ), but not so good in using "mathematical reasoning skills" ( $A_{10}$ ), "sorting and setting goals in multistep problems" ( $A_{14}$ ), and weak in using "test-taking skills" ( $A_{13}$ ).

In short, the students in Group A tend to be excellent in using process-oriented attributes, and they know more mathematics than the students in the other two groups. The members of Group C tend to be good in application-oriented attributes, but not process-oriented attributes. They have better language skills than the members of the other two groups, and they may be good in arithmetic, but they are not good in either algebra or geometry. Group B is the opposite of Group A in terms of canonical discriminant function 1, but very similar to Group A in terms of canonical discriminant function 2.

### Discriminant Analysis for Racial-Ethnic Groups

Discriminant analysis was carried out for the four racial-ethnic groups: Asians, Blacks, Hispanics, and Whites. Table 29.8 gives group-level information on the frequencies in the four groups and the Mahalanobis distances between the groups. The distance between the Asian and the Black groups is the largest, followed by the distance between the Asian and the Hispanic groups; but the distances between the White and Hispanic, and the Black and Hispanic groups are smaller. Discriminant analysis optimizes these group differences with respect to the attribute-probability scores. Table 29.8 summarizes the discriminant analysis.

TABLE 29.8.  
Squared Mahalanobis Distances Between Four Racial-Ethnic Groups and Their  
Frequencies

| Group    | Frequency | Asian | Black | Hispanic |
|----------|-----------|-------|-------|----------|
| Asian    | 236       |       |       |          |
| Black    | 369       | 2.92  |       |          |
| Hispanic | 219       | 1.66  | 0.57  |          |
| White    | 2,819     | 0.64  | 1.33  | 0.35     |

TABLE 29.9.

Standardized Canonical Coefficients and Group Means for the Four Racial-Ethnic Groups

| Attributes                                           | Canonical Coefficient 1 | Canonical Coefficient 2 | Canonical Coefficient 3 |
|------------------------------------------------------|-------------------------|-------------------------|-------------------------|
| A <sub>1</sub> (Arithmetic)                          | 0.246                   | -0.273                  | 0.356                   |
| A <sub>2</sub> (Elementary algebra)                  | 0.088                   | 0.062                   | -0.197                  |
| A <sub>3</sub> (Intermediate algebra)                | 0.163                   | 0.215                   | -0.162                  |
| A <sub>4</sub> (Geometry)                            | 0.173                   | -0.116                  | -0.288                  |
| A <sub>5</sub> (Word problem)                        | 0.320                   | -0.230                  | 0.341                   |
| A <sub>6</sub> (Comparison format)                   | 0.240                   | -0.121                  | 0.176                   |
| A <sub>7</sub> (Application of knowledge)            | 0.219                   | 0.140                   | 0.266                   |
| A <sub>8</sub> (Application of rules and algorithms) | 0.147                   | -0.070                  | 0.017                   |
| A <sub>9</sub> (Selection of rules and theorems)     | 0.013                   | 0.066                   | -0.343                  |
| A <sub>10</sub> (Reasoning and logical thinking)     | 0.204                   | 0.464                   | -0.405                  |
| A <sub>11</sub> (Analytical thinking)                | 0.107                   | 0.298                   | 0.284                   |
| A <sub>12</sub> (Working with figures and tables)    | 0.440                   | -0.402                  | -0.365                  |
| A <sub>13</sub> (Test-taking skills)                 | -0.083                  | 0.303                   | 0.268                   |
| A <sub>14</sub> (Sorting subgoals)                   | 0.223                   | 0.393                   | 0.357                   |
| Asian                                                | 0.665                   | 0.512                   | -0.070                  |
| Black                                                | -1.014                  | 0.202                   | 0.041                   |
| Hispanic                                             | -0.420                  | -0.166                  | -0.227                  |
| White                                                | 0.110                   | -0.057                  | 0.018                   |

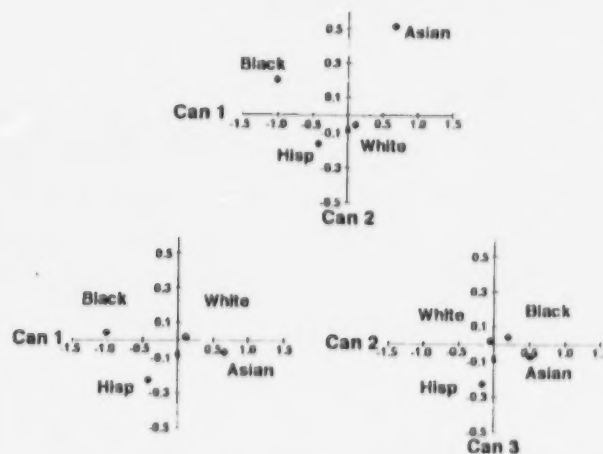


FIG. 29.14. Class means for the four racial-ethnic groups.

The racial-ethnic groups are fairly well separated, and they are different with respect to their attributes. Group differences can be seen clearly in FIG. 29.14, which gives plots of the

four centroids with respect to the canonical discriminant functions described in Table 29.9. In Table 29.9, the first canonical discriminant function has large coefficients for Attributes 12 (following verbal instructions, comprehending tables, figures, and terminologies), 5 (word problems), and 1 (basic knowledge and skills in arithmetic). The dimension defined by the first canonical discriminant function denotes examinees having a better foundation in mathematics and being able to do the problems associated with verbal competency.

The second canonical discriminant function has large coefficients for Attributes 10 (mathematical reasoning skills), 14 (sorting the goals in multistep problems), 13 (test-taking skills), and 11 (analytical thinking skills) in the positive direction, and for Attributes 12 and 1 in the negative direction. The second dimension measures competency in mathematical thinking skills, including test-taking skills. The third canonical discriminant function is characterized by Attributes 1, 5, 14, 11, and 7 in the positive direction, and Attributes 4, 9, 10, and 12 in the negative direction.

The summary of the group descriptions shows that Asian and White students are at the positive end of the first dimension, but Black and Hispanic students are not. However, Black students are doing as well as White students in the second dimension, whereas Asian students obtain the highest positive mean. It should be noted that the mean of  $A_1$  for Black students is as high as that for White students, and the differences of the means for  $A_{10}$ ,  $A_{11}$ , and  $A_{14}$  among White, Black, and Hispanic students are very close, less than 1 SD. Hispanic students have a tendency to have negative values for the third canonical discriminant function. Unfortunately, it is difficult to interpret this third function.

### Discriminant Analysis for the Gender Differences

The female and male groups are investigated by discriminant analysis too. Table 29.10 summarizes their group-level information, frequencies, and Mahalanobis distances between the groups. The total-sample standardized canonical coefficients for the gender groups are given in Table 29.11. FIG. 29.15 shows the plot of the canonical means for these groups.

TABLE 29.10.  
Squared Mahalanobis Distances Between the Two Gender Groups and Their Frequencies

| <i>Gender</i> | <i>Frequency</i> | <i>Female</i> |
|---------------|------------------|---------------|
| Female        | 3,916            |               |
| Male          | 2,647            | 0.45          |

TABLE 29.11

Total-Sample, Standardized Canonical Coefficients and Group Means for the Two Gender Groups

| Attributes                                           | Canonical Coefficient 1 |
|------------------------------------------------------|-------------------------|
| A <sub>1</sub> (Arithmetic)                          | 0.038                   |
| A <sub>2</sub> (Elementary algebra)                  | 0.071                   |
| A <sub>3</sub> (Intermediate algebra)                | 0.128                   |
| A <sub>4</sub> (Geometry)                            | 0.039                   |
| A <sub>5</sub> (Word problems)                       | 0.398                   |
| A <sub>6</sub> (Comparison format)                   | 0.034                   |
| A <sub>7</sub> (Application of knowledge)            | 0.361                   |
| A <sub>8</sub> (Application of rules and algorithms) | -0.054                  |
| A <sub>9</sub> (Selection of rules and theorems)     | -0.001                  |
| A <sub>10</sub> (Reasoning and logical thinking)     | 0.331                   |
| A <sub>11</sub> (Analytical thinking)                | 0.217                   |
| A <sub>12</sub> (Working with figures and tables )   | 0.158                   |
| A <sub>13</sub> (Test-taking skills)                 | 0.289                   |
| A <sub>14</sub> (Sorting subgoals)                   | 0.355                   |
| Female                                               | -0.270                  |
| Male                                                 | 0.400                   |

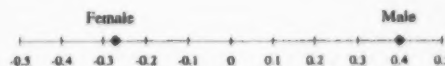


FIG. 29.15. Class means for the two gender groups.

With regard to the gender differences (FIG. 29.15), men perform better than women on Attributes 5, 7, 10, 13, and 14. Men tend to be more knowledgeable, stronger in translating word problems, and to have better mathematical reasoning skills and have better test-taking skills than women. However, the differences at the attribute levels are much smaller in relation to their standard errors than the total score levels or IRT  $\theta$  levels, and the mean differences for the attribute scores are not statistically significant.

## CONCLUSIONS

This study showed that an analysis of the attributes required to answer correctly items in Section I of the GRE-Q utilizing the rule-space methodology is useful in answering several educational and psychometric questions:

- Predicting the difficulty of items from a Q-matrix—a matrix representing item-and-attribute involvement relationships—is useful in tapping the underlying cognitive processes and knowledge that are required to solve the problems, especially when the number of items is substantially larger than the number of attributes. The matrix indicates clearly that difficult items involve more mathematical thinking attributes than easy items do. The content attributes are often statistically weak in predicting the difficulty of items.
- Attribute-probability scores are a better predictor of IRT  $\theta$  values than attribute-mastery patterns. Because, in this study, the number of attributes is 14 and the number of items is 30, a multiple  $R^2$  of around .80 with 14 variables is very good when  $R^2 = .89$  with 30 variables.
- For gender and racial-ethnic groups, the differences in attribute means are much smaller in relation to their standard errors than the differences in the means of  $\theta$ s or in the total scores. The differences in the attribute means are smaller than a standard deviation, whereas the differences in the means of  $\theta$ s or the means of scale scores are larger than a standard deviation. It is obvious that we have an “amplified effect,” that is, the accumulation of small differences in each of the 14 attribute scores becomes a large difference.
- For racial-ethnic groups, the content attributes and the attributes involving verbal comprehension have much larger mean differences in relation to their standard errors than thinking attributes and application attributes have. Some thinking attributes (such as Attribute 13) that may not require cumulative training have the same mean value for Black and White students.
- Cluster analyses of the 33 majors, 4 racial-ethnic groups, and 2 gender groups were performed, and three main groups were identified. The first group (Group A) consists of Asian students, and those majoring in civil engineering, computer science, economics, chemical engineering, industrial engineering, mechanical engineering, physics, electrical engineering, and mathematics. The second group (Group B) is

characterized by Black and Hispanic students and those majoring in social work, elementary education, home economics, special education, child education, communications, sociology, medicine, performing art, and secondary education. The third group (Group C) is characterized by White students, men and women, and students majoring in agriculture, anthropology, history, political science, art history, foreign languages, business administration, English, psychology, architecture, biology, and earth-atmospheric-marine science.

- A discriminant analysis characterized three clustered groups: Group A is associated with an excellent use of processing skills and knows more mathematics than the other two groups. Group B is associated with better application skills, but not better processing skills. Group B is the opposite of Group A with respect to the first canonical coordinates, but they are about same with respect to the second canonical coordinates. Group C is identified by large second canonical coordinates; that is, members of Group C are better in arithmetic, have a better knowledge of middle-school-level mathematics in general, and are excellent in application skills. They are more likely to lack knowledge and basic skills in higher level algebra and mathematics and are naive in using mathematical thinking skills.
- For racial-ethnic groups, discriminant analysis revealed Hispanics and Blacks tend to be considerably weaker in basic knowledge and skills in arithmetic, elementary algebra, and geometry than Asians and Whites. However, they are less weak in terms of process attributes.
- Concerning gender differences, men tend to be more knowledgeable, to be stronger in solving word problems, and to have better mathematical reasoning skills and better test-taking skills than women do. However, the differences in relation to their standard errors at the attribute level are much smaller than at the total score level or at the IRT  $\theta$  level. The mean differences for the attribute scores are not statistically significant. In Group C, women cluster with English, psychology, and business

administration majors, whereas men cluster with architecture, biology, and earth-atmospheric-marine-majors.

## DISCUSSION

An incidence matrix  $Q$  of attributes and items, whose elements indicate the involvement or noninvolvement of each attribute in the items, prompts new item-analysis techniques associated closely with attribute statistics. Rule-space methodology enables us to calculate attribute difficulties and discriminating powers. By knowing the difficulties and the values of point-biserials for the attributes and total scores for the attributes and IRT  $\theta$ , and by knowing which attributes are involved in an item, one should be able to estimate the item difficulties and item-discriminating powers from the difficulties and the discriminating powers of attributes. There should be a mathematical relationship between the reliability of the total score by items and the total scores by attributes.

Test developers can select a set of attributes that they would like to test by looking into the values of attribute difficulties and discrimination powers. By selecting the appropriate attributes, the difficulty of items can be manipulated. If this relationship were developed fully, then the current practice of item development can be improved, and the costs incurred in developing tests can be substantially reduced. Most of all, the costs of administering pretests will be much reduced. However, new psychometric theories will need to be developed.

This chapter shows that there are subgroup differences in terms of performances on the attributes. Hence it becomes easier to tailor item difficulties to a particular population by selecting an appropriate set of attributes, such as the use of analytical thinking skills. The developers of quantitative tests for science and engineering majors can ignore easy attributes such as  $A_1$ ,  $A_2$ ,  $A_4$ , and  $A_{12}$  when constructing tests. What science and engineering students have to know are more difficult mathematical thinking skills interacting with knowledge of higher level mathematics than those tested in the current GRE-Q tests.

The development of a fair test for all of the subgroups is very important. To achieve this goal, we must know the sources of the differences between the groups. Studying protocols has been one of the most useful techniques for investigating the cognitive aspect of items. If taking protocols and/or interviewing were done in conjunction with using these statistical techniques, then the shortcomings associated with either of the two approaches would be

compensated for by the other. Statistical analysis is objective, but it lacks substantive information. Protocol analyses can give us qualitative and descriptive information, which may not be objective enough. In other words, the rule-space methodology can be used together with protocol analyses.

The methodology used in this study has opened up various possibilities for the future such as the development of new test theories that can be useful for generating new items associated with attributes, the diagnosis of students' latent knowledge states and the use of that information for improving learning and teaching (K. K. Tatsuoka & C. M. Tatsuoka, 1998), and the development of conditional probability curves for attributes and the use of the curves for proficiency scaling (K. K. Tatsuoka, et al., 1994).

Chipman and Nichols (1995) suggested that the rule-space methodology should be helpful in creating cognitive models. Researchers have already used it for guiding the development of their cognitive models in such complex domains as architecture (I. R. Katz, Martinez, Sheehan, & Tatsuoka, 1994), algebra (Birenbaum, et al., 1993) and the GRE-Q (K. K. Tatsuoka & Gallagher, in press).

### ACKNOWLEDGMENTS

The authors gratefully acknowledge and thank Maurice Tatsuoka who gave much editorial help and statistical consultation, Curtis Tatsuoka who engaged in many useful discussions with the authors, Phil Leung for data analyses and graphic design, and Liz Brophy for preparing the manuscript.

## APPENDIX A

### Major Fields of Study

#### *Humanities and Art*

- Architecture
- Art history
- English
- Foreign languages
- History
- Performing arts
- Philosophy

#### *Natural Sciences*

- Agriculture
- Biology
- Home economics
- Medicine

#### *Physical Sciences: Nonengineering*

- Chemistry
- Computer science
- Earth, atmospheric, and marine sciences
- Mathematics
- Physics

#### *Physical Sciences: Engineering*

- Chemical engineering
- Civil engineering
- Electrical engineering

- Industrial engineering
- Mechanical engineering

*Social Sciences and Business*

- Anthropology
- Business administration
- Communication
- Economics
- Political science
- Psychology
- Sociology

*Social Work and Education*

- Social work
- Child education
- Elementary education
- Secondary education
- Special education

## APPENDIX B

### Attributes Required by Examinees to Answer Correctly Items on the GRE Quantitative Test

*Content Variables*

$A_1$  (*Arithmetic—Numbers and Operations*). Can work with odd and even integers, prime numbers, factors, rational numbers, ordering, ratios, percentages, square of positive numbers, square root of small positive numbers such as 4, 9, 16, average of positive numbers.

$A_2$  (*Elementary Algebra*). Can work with variables (addition and subtraction only), linear equations, linear algebraic expressions, signed numbers, absolute values, irrational numbers (such as the square root of 3).

*A<sub>1</sub> (Intermediate Algebra).* Can work with factorable quadratic equations, complicated algebraic expressions, functions, sets, simple probability, simple combinations, exponents with variables, simple statistics such as modes and medians.

*A<sub>2</sub> (Geometry and Analytic Geometry).* Can work with perimeter, area, and volume of triangles, circles, rectangles, and other geometric figures. In analytic geometry, can work with points, lines, and their expressions in terms of a coordinate system.

#### *Process Variables*

*A<sub>3</sub> (Translating Word Problems).* Can translate a word problem into an arithmetic expression(s), algebraic expression(s), equation(s). Can identify implicit variables and relationships.

*A<sub>4</sub> (Application of Knowledge).* Can recall and interpret knowledge about basic concepts and properties in arithmetic, elementary algebra, intermediate algebra, geometry, and analytic geometry. Can perform computations in arithmetic (ratio, powers), geometry, signed numbers, absolute values, median, and mode.

*A<sub>5</sub> (Application of Rules and Algorithms).* Can solve equations, solve simultaneous linear equations, derive algebraic expressions, compute algebraic expressions, factor algebraic equations, factor algebraic expressions, substitute numbers and/or variables in algebraic expressions and algebraic equations.

*A<sub>6</sub> (Selection of Rules and Theorems).* Can select and apply to the task the relevant concept, theorem, rule, and/or strategy from a large knowledge base.

*A<sub>10</sub> (Reasoning and Logical Thinking).* Can reason from cause to effect, identify and understand necessary and sufficient conditions and apply them appropriately to solve problems, check that solutions hold true in the general case as opposed to a specific case.

*A<sub>11</sub> (Analytical Thinking and Problem Structuring).* Can analyze the problem into component parts or constituent elements, which are implicit, and structure the component parts so that a problem becomes solvable.

*A<sub>13</sub> (Test-Taking Skills).* Can take advantage of the form of the test items and other test-taking methods without solving the problem in the manner intended by the item writer. Can solve a task by working back from the multiple-choice options, producing one or two examples and, from them, inferring the correct option, substituting numbers for selecting the

correct option, eliminating one or more of the options, and guessing from among the remaining options.

*A<sub>14</sub> (Degree of Complexity).* Can deal with problems that involve several steps. These steps may be explicit or implicit. Can establish the subgoals of the problem, organize them in relation to each other, rate them in order of priority, and execute them in the proper sequence in a step-by-step fashion.

*Context Variables*

*A<sub>6</sub> (Comparison Format).* The item is written in a comparison format.

*A<sub>12</sub> (Working With Figures, Tables and Graphs).* Can comprehend tables, figures, and graphs. Can follow verbally described mathematical problems. In the statement of the problem, the student can relate the geometric terms with figures, graphs, and charts. Can generate figures for using external representations to facilitate problem-solving activities.



*U.S. Department of Education  
Office of Educational Research and Improvement (OERI)  
National Library of Education (NLE)  
Educational Resources Information Center (ERIC)*

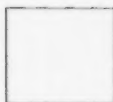


## **NOTICE**

### **Reproduction Basis**



This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.



This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").